

20
CLAIMS

1. A method comprising:

- 5 providing a distributed switch fabric network having a first node and a second node coupled to exchange a plurality of packets over a channel, wherein the plurality of packets are divided into a plurality of priority levels of packets;
- detecting a congestion condition of one of the plurality of priority levels of packets in the second node;
- 10 reporting the congestion condition of the one of the plurality of priority levels of packets to the first node;
- the first node suspending transmission of the one of the plurality of priority levels of packets over the channel to the second node;
- detecting a clear condition of the one of the plurality of priority levels of packets in the second node;
- 15 reporting the clear condition of the one of the plurality of priority levels of packets to the first node; and
- resuming transmission of the one of the plurality of priority levels of packets from the first node to the second node.

20 2. The method of claim 1, further comprising:

- providing each of the plurality of priority levels of packets with a transmit buffer;
- the one of the plurality of priority levels of packets accumulating in the transmit buffer corresponding to the one of the plurality of priority levels of packets;
- 25 if the transmit buffer corresponding to the one of the plurality of priority levels of packets reaches a transmit threshold value, a traffic manager of the first node modifying transmission of the one of the plurality of priority levels of packets to the transmit buffer;
- and
- if resuming transmission occurs prior to the transmit buffer reaching the transmit threshold value, the first node suspending transmission of the one of the plurality of
- 30 priority levels of packets occurring transparently to the traffic manager of the first node.

3. The method of claim 2, wherein the first node modifying transmission of the one of the plurality of priority levels of packets comprises suspending transmission of the one of the plurality of priority levels of packets over the channel to the transmit buffer.

5 4. The method of claim 2, wherein modifying transmission of the one of the plurality of priority levels of packets comprises throttling transmission of the one of the plurality of priority levels of packets over the channel to the transmit buffer.

10 5. The method of claim 1, further comprising allowing the plurality of packets other than the one of the plurality of priority levels of packets to continue to the second node over the channel.

15 6. The method of claim 1, wherein first node comprises a first node transceiver port and the second node comprises a second node transceiver port, and wherein detecting the congestion condition comprises the second node transceiver port detecting the congestion condition.

20 7. The method of claim 6, wherein reporting the congestion condition comprises the second node transceiver port reporting the congestion condition to the first node transceiver port.

 8. The method of claim 6, wherein detecting the clear condition comprises the second node transceiver port detecting the clear condition.

25 9. The method of claim 6, wherein reporting the clear condition comprises the second node transceiver port reporting the clear condition to the first node transceiver port.

 10. A method comprising:
 providing a distributed switch fabric network having a first node and a second
30 node coupled to communicate over a channel, wherein at least the first node distinguishes between a plurality of priority levels of packets;
 if the second node distinguishes between the plurality of priority levels of packets:

the first node suspending transmission of one of the plurality of priority levels of packets over the channel to the second node upon a congestion condition of the one of the plurality of priority levels of packets at the second node;

5 the first node resuming transmission of the one of the plurality of priority levels of packets over the channel to the second node upon a clear condition of the one of the plurality of priority levels of packets at the second node; and
if the second node fails to distinguish between the plurality of priority levels of packets:

10 the first node suspending transmission of all of the plurality priority levels of packets over the channel to the second node upon the congestion condition at the second node;

the first node resuming transmission of all of the plurality of priority levels of packets over the channel to the second node upon the clear condition at the second node.

15

11. The method of claim 10, wherein if the second node distinguishes between a plurality of priority levels of packets, the second node comprising at least one of:

a plurality of transmit buffers, wherein each of the plurality of transmit buffers corresponds to one of the plurality of priority levels of packets; and

20 a plurality of receive buffers, wherein each of the plurality of receive buffers corresponds to one of the plurality of priority levels of packets.

12. A method comprising:

25 providing a distributed switch fabric network having a first node having a first node transceiver port and a second node having a second node transceiver port;

link level flow control operating between the first node transceiver port and the second node transceiver port to in response to a congestion condition in the second node transceiver port, wherein the link level flow control suspends transmission of one of a plurality of priority levels of packets on a channel from the first node transceiver port to
30 the second node transceiver port;

the one of the plurality of priority levels of packets accumulating in one of a plurality of transmit buffers of the first node transceiver port, wherein the one of the

plurality of transmit buffers corresponds to the one of the plurality of priority levels of packets;

per-flow flow control operating to modify transmission of the one of the plurality of priority levels of packets to the one of the plurality of transmit buffers if the one of the plurality of transmit buffers reaches a transmit threshold value; and

link level flow control operating transparently to a traffic manager of the first node if the congestion condition occurs and the one of the plurality of transmit buffers fails to reach the transmit threshold value.

10 13. The method of claim 12, wherein the first node transceiver port is comprised of a first node receiver port and a first node transmitter port, wherein the second node transceiver port is comprised of a second node receiver port and a second node transmitter port, and wherein the link level flow control operating comprises:

the second node receiver port detecting the congestion condition;

15 the second node receiver port reporting the congestion condition to the second node transmitter port;

the second node transmitter port reporting the congestion condition to the first node receiver port;

20 the first node receiver port transmitting a priority level stop signal to the first node transmitter port, wherein the priority level stop signal corresponds to the one of the plurality of priority levels of packets; and

the first node transmitter port suspending transmission of the one of the plurality of priority levels of packets to the second node receiver port.

25 14. The method of claim 12, wherein the first node transceiver port is comprised of a first node receiver port and a first node transmitter port, wherein the second node transceiver port is comprised of a second node receiver port and a second node transmitter port, and wherein the link level flow control operating comprises:

the second node receiver port detecting a clear condition;

30 the second node receiver port reporting the clear condition to the second node transmitter port;

the second node transmitter port reporting the clear condition to the first node receiver port;

the first node receiver port transmitting a priority level start signal to the first node transmitter port, wherein the priority level start signal corresponds to the one of the plurality of priority levels of packets; and

the first node transmitter port resuming transmission of the one of the plurality of priority levels of packets to the second node receiver port.

15. The method of claim 12, wherein the per-flow flow control operating comprises:

a flow control generator detecting the one of the plurality of transmit buffers reaching the transmit threshold value and transmitting a modify flow control packet to the traffic manager, wherein the modify flow control packet operates to modify the flow of the one of the plurality of priority levels of packets to the one of the plurality of transmit buffers; and

the flow control generator detecting the one of the plurality of transmit buffers dropping below the transmit threshold value and transmitting a resume transmission packet to the traffic manager, wherein the resume transmission packet operates to resume transmission of the one of the plurality of priority levels of packets to the one of the plurality of transmit buffers.

16. The method of claim 15, further comprising the flow control generator scheduling transmission of a plurality of priority levels of packets, the modify flow control packet and the resume transmission packet to the traffic manager.

17. The method of claim 16, wherein the flow control generator giving priority to the modify flow control packet and the resume transmission packet over the plurality of priority levels of packets.

18. The method of claim 12, wherein operating to modify transmission of the one of the plurality of priority levels of packets comprises suspending transmission of the one of the plurality of priority levels of packets over the channel to the one of the plurality of transmit buffers.

19. The method of claim 12, wherein operating to modify transmission of the one of the plurality of priority levels of packets comprises throttling transmission of the one of the plurality of priority levels of packets over the channel to the one of the plurality of transmit buffers.

5

20. A method comprising:

providing a first node, wherein the first node is coupled to exchange a plurality of packets with a second node over a channel in a distributed switch fabric network, wherein the plurality of packets are divided into a plurality of priority levels of packets;

10 the first node generating the plurality of priority levels of packets for transmission to the second node over the channel;

separating each of the plurality of priority levels of packets into each of a plurality of transmit buffers, wherein each of the plurality of transmit buffers corresponds to one of the plurality of priority levels of packets; and

15 scheduling transmission of the plurality of priority levels of packets to the second node over the channel, wherein scheduling selects which of the plurality of priority levels of packets from each of the plurality of transmit buffers are transmitted to the second node.

20 21. The method of claim 20, wherein the separating and the scheduling occurring separately at each first node transceiver port.

22. The method of claim 20, wherein if the second node distinguishes between the plurality of priority levels of packets:

25 the first node suspending transmission of one of the plurality of priority levels of packets over the channel to the second node upon a congestion condition of the one of the plurality of priority levels of packets at the second node; and

the first node resuming transmission of the one of the plurality of priority levels of packets over the channel to the second node upon a clear condition of the one of the
30 plurality of priority levels of packets at the second node.

23. The method of claim 20, wherein if the second node fails to distinguish between the plurality of priority levels of packets:

the first node suspending transmission of all of the plurality priority levels of packets over the channel to the second node upon the congestion condition at the second node; and

5 the first node resuming transmission of all of the plurality of priority levels of packets over the channel to the second node upon a clear condition at the second node.

24. The method of claim 20, wherein if the second node distinguishes between a plurality of priority levels of packets, the second node comprising a plurality of receive buffers, wherein each of the plurality of receive buffers corresponds to one of the plurality
10 of priority levels of packets.

25. A method comprising:

providing a first node, wherein the first node is coupled to exchange a plurality of packets with a second node over a channel in a distributed switch fabric network, wherein
15 the plurality of packets are divided into a plurality of priority levels of packets;

the first node receiving the plurality of priority levels of packets;

separating each of the plurality of priority levels of packets into each of a plurality of receive buffers, wherein each of the plurality of receive buffers corresponds to one of the plurality of priority levels of packets; and

20 scheduling processing of the plurality of priority levels of packets by the first node, wherein scheduling selects which of the plurality of priority levels of packets from each of the plurality of receive buffers are processed by the first node.

26. The method of claim 25, wherein the separating occurring separately at each
25 first node transceiver port.

27. The method of claim 25, wherein scheduling comprises selecting the plurality of priority levels of packets from any of the plurality of receive buffers of each first node transceiver port.

30

28. The method of claim 25, wherein if the second node distinguishes between the plurality of priority levels of packets:

the second node suspending transmission of one of the plurality of priority levels of packets over the channel to the first node upon a congestion condition of the one of the plurality of priority levels of packets at the first node; and

the second node resuming transmission of the one of the plurality of priority levels of packets over the channel to the first node upon a clear condition of the one of the plurality of priority levels of packets at the first node.

29. The method of claim 25, wherein if the second node fails to distinguish between the plurality of priority levels of packets:

the second node suspending transmission of all of the plurality priority levels of packets over the channel to the first node upon the congestion condition at the first node; and

the second node resuming transmission of all of the plurality of priority levels of packets over the channel to the first node upon a clear condition at the first node.

30. The method of claim 25, wherein if the second node distinguishes between a plurality of priority levels of packets, the second node comprising a plurality of transmit buffers, wherein each of the plurality of transmit buffers corresponds to one of the plurality of priority levels of packets.

31. A method comprising:

providing a first node having a first node receiver port, wherein the first node receive port comprises a plurality of receive buffers,

the first node receiver port receiving a plurality of packets from a second node over a channel in a distributed switch fabric network, wherein the plurality of packets are divided into a plurality of priority levels of packets, and wherein each of the plurality of receive buffers corresponds to one of the plurality of priority levels of packets;

detecting a congestion condition of one of the plurality of priority levels of packets in the second node;

if the congestion condition is for a highest priority level of packets the second node suspending transmission of all of the plurality of priority levels of packets to the first node receiver port, and wherein if the congestion condition is not for the highest priority level

of packets the second node suspending transmission of the one of the plurality of priority levels of packets to the first node;

detecting a clear condition of the one of the plurality of priority levels of packets in the second node; and

- 5 resuming transmission of the one of the plurality of priority levels of packets to the first node.